

Reconhecimento de imagens de gestos e poses de mão em um ambiente automotivo

Gustavo Müller Nunes

January 2014

Sumário

1	Introdução	4
1.1	Apresentação	4
1.2	Objetivo	4
1.3	Justificativa	5
1.4	Hipótese	5
1.5	Metodologia	5
1.6	Organização da dissertação	6
2	Revisão de Literatura	7
2.1	Momento	7
2.2	Momentos invariantes em translação, rotação e escala	7
2.2.1	Introdução	7
3	Material e método	8
3.1	Materiais	8
3.2	Sistema Proposto	8
3.3	Discussão	8
3.3.1	Câmera IR	8
3.4	Construção da câmera IR	8
4	Resultados e discussão	10
5	Conclusão	11

Lista de Tabelas

Lista de Figuras

1.1	Kinect, da Microsoft, e a câmera da <i>Creative</i> com parceria da Intel	4
3.1	Webcam sem modificações	8
3.2	Webcam sem modificações	9
3.3	Webcam sem modificações	9
3.4	Webcam sem modificações	9
3.5	Webcam sem modificações	9

Capítulo 1

Introdução

1.1 Apresentação

Reconhecimento de gestos baseado em visão computacional é um assunto bastante pesquisado e já pode ser considerado popular, isto porque, a busca por mecanismos que tornem a interação entre homem e máquina mais intuitiva e natural é constante e vem aumentando com o lançamento de plataformas que auxiliam os desenvolvedores nos complexos algoritmos que envolvem essa área. O lançamento do Kinect, da Microsoft [3], e da plataforma de desenvolvimento da Intel, chamada Intel Perceptual Computing [4] (ambas com câmeras de profundidade) vem popularizando o desenvolvimento de aplicativos e revolucionando o jeito que interagimos com os jogos e computadores.

NOTA: Contextualizar a figura



Figura 1.1: Kinect, da Microsoft, e a câmera da *Creative* com parceria da Intel

O uso de câmeras em carros e caminhões também tem aumentando nos últimos anos. Sistemas de segurança capazes de verificar se o motorista está saindo indevidamente da faixa, ou se o veículo está em rota de colisão com algum outro automóvel ou objeto e até mesmo monitorando o stress do motorista já são comuns em vários modelos de veículos. Mas pouco vimos o uso dessas câmeras para interação do motorista com a grande quantidade de controles que temos nos carros. Aumentar ou diminuir o volume do rádio, trocar de faixa de música, dar zoom no mapa do sistema de navegação são alguns exemplos de comandos que poderia ser dados por meio de gestos. O sistema de gestos também pode ser usado como um complemento ao sistema de reconhecimento de voz, bastante comum hoje nos carros e que funciona muito bem.

1.2 Objetivo

O objetivo do trabalho é discutir as principais técnicas para reconhecimento de gestos e poses de mão em um ambiente automotivo. Os algoritmos e metodologias hoje utilizados para segmentar e extrair características de imagens e vídeos devem ser estudados e verificados se atingem seu propósito em um

ambiente automotivo. As características extraídas são utilizadas como entrada em um classificador, responsável por reconhecer gestos e poses de mão, e assim, permitir uma interação com o veículo traduzindo os gestos em comandos para o carro.

1.3 Justificativa

Esse ambiente apresenta uma forte variação de luz e ausência de controle nas características da mão e do braço do motorista (cor de pele, braço com ou sem vestimentas e vestimentas de cores e estampas diferentes).

As condições gerais dentro do automóvel inclui uma grande variação de iluminação, mudança de usuário e fundos não uniformes. Além disso, a aceitação do usuário é um item bastante importante, portanto coisas como uma iluminação artificial visível, restrição de vestimentas e calibração extensiva não pode ser tolerados. Tendo isso em mente, alguns critérios e requisitos para o sistema podem ser estabelecidos:

- robustez contra ambientes ruidosos
- iluminação invisível
- independente de usuário
- sem calibração ou treinamento pelo usuário
- pequeno e compreensível conjunto de gestos
- reação do sistema com o mínimo de latência

1.4 Hipótese

As metodologias apresentadas para a segmentação, extração de características e classificação precisam ser estudadas e implementadas para verificar se o processamento requerido viabiliza o seu uso em uma área onde o custo é relativamente sensível.

1.5 Metodologia

A metodologia do projeto abrange todas as principais etapas de um problema de visão computacional, com a captura da imagem, segmentação, extração de características, reconhecimento de padrões e interpretação dos resultados obtidos.

As imagens de poses e os vídeos dos gestos serão obtidos através de uma câmera infravermelha, conforme xxxx, em dois ambientes distintos. Primeiro em um ambiente controlado com fundo homogêneo de cor preta e em uma sala totalmente escura (essa base de dados será usada como referência para os algoritmos implementados). O outro será obtido no interior de um veículo, tanto de dia como de noite.

A segmentação da imagem é feita sem o auxílio de luvas ou marcadores. Essa etapa é responsável por melhor dividir as regiões da imagem para que a etapa de extração de características possa avaliar qual é a mais provável de ser a mão.

As características extraídas variam conforme o método de classificação ou se o reconhecimento é de pose ou gesto. Serão estudados os principais algoritmos para o mesmo e verificar o que melhor atende o objetivo.

Os classificadores mais utilizados na literatura existente serão avaliados, para que se conheça sua performance em relação a tempo de processamento e acerto. Desse estudo serão terminados os classificadores mais adequados para a aplicação proposta.

1.6 Organização da dissertação

NOTA: <i>Elaborar no final</i>

Capítulo 2

Revisão de Literatura

2.1 Momentos

Momentos são medições escalares usadas para caracterizar uma função e capturar suas características mais significativas. São bastante usados a centenas de anos em estatística para descrever a forma de uma função de densidade probabilística e em corpos rígidos para medir a distribuição de massa. Do ponto de vista matemático, momentos são "projeções" de uma função em uma base polinomial (da mesma forma que, transformada de Fourier é uma projeção em uma base de funções harmônicas).

Definindo então uma imagem como sendo uma função real $f(x, y)$ de duas variáveis em a compact support $D \subset \mathbb{R} \times \mathbb{R}$ e tendo uma integral finita diferente de zero. Podemos definir o momento de forma genérica $M_{pq}^{(f)}$ de uma imagem $f(x, y)$, onde p, q são valores não negativos e inteiros e $r = p + q$ é a ordem do momento, como:

$$M_{pq}^{(f)} = \iint_D p_{pq}(x, y) f(x, y) dx dy$$

2.2 Momentos invariantes em translação, rotação e escala

2.2.1 Introdução

Translação, rotação e escala (abreviado como TRS, do inglês *Translation, rotation and scaling*) são as transformações de coordenadas espacial mais simples. TRS é uma transformada de 4 parâmetros, que pode ser descrita como

$$x' = sR \cdot x + t$$

NOTA: Verificar link http://docs.opencv.org/doc/tutorials/imgproc/shapedescriptors/moments/moments.html
--

Capítulo 3

Material e método

3.1 Materiais

3.2 Sistema Proposto

3.3 Discussão

3.3.1 Câmera IR

NOTA: Escrever um pouco sobre as câmeras IR

3.4 Construção da câmera IR

NOTA: Um pouco de texto



Figura 3.1: Webcam sem modificações

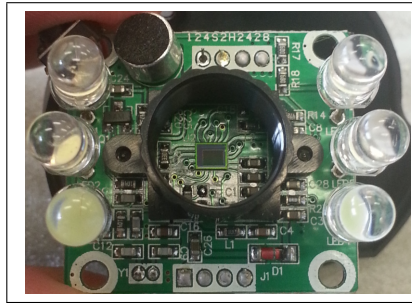


Figura 3.2: Webcam sem modificações

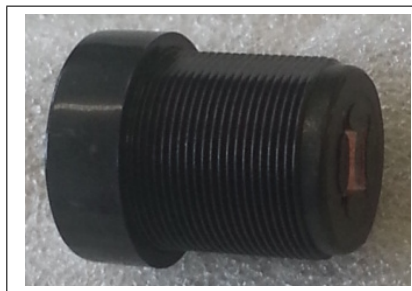


Figura 3.3: Webcam sem modificações



Figura 3.4: Webcam sem modificações



Figura 3.5: Webcam sem modificações

Capítulo 4

Resultados e discussão

Capítulo 5

Conclusão

Referências Bibliográficas

- [1] Zobl, M., Nieschulz, R., Geiger, M., Lang M., Rigoll, G., Gesture Components for Natural Interaction with In-Car devices, 2003.
- [2] Akyol, S., Canzler, U., Bengler, K., Hahn, W.: Gesture control for use in auto-mobiles. In: Proceedings, MV A 2000 Workshop on Machine Vision Applications, Tokyo, Japan, November 28-30, 2000, IAPR, ISBN 4-901122-00-2 (2000)
- [3] <http://www.microsoft.com/en-us/kinectforwindows/develop/>
- [4] <http://software.intel.com/en-us/vcsourcetoold/perceptual-computing-sdk>